

## 明細書

### 音声合成装置

### 技術分野

[0001] 本発明は、音声合成装置に関し、詳しくは、事前に収録された所定文章の音声データを所定の音声単位で複数蓄積した音声データベースを搭載し、任意のテキストデータに対して前記音声データベースに基づきコーパスベース音声合成を行うための、他の機器に組込み搭載される組込みマイコンからなる音声合成装置に係わる。

### 背景技術

[0002] 従来、音声合成技術として、駅の案内放送等に利用されるような、予め、使用される所定の単語やフレーズを音源として収録しておき、これらを組み合わせることで、機械による文章の音声出力を行う録音編集方式と、電話の自動案内等に利用されるような、事前に一文字ずつ音声波形に近い音データを蓄積し、これらの音データを信号処理により単音接続し、肉声の音声波形に近い音として出力する規則合成方式があった。

[0003] ここで、録音編集方式では、事前に収録された決められたフレーズによる録音データの組み合わせのみが可能であることから、合成可能な文章数が限定されてしまい、また、新たな文章を追加して合成させたいときには、この追加文章に使用される単語やフレーズの音源の再録音が必要となるために、再録音のための費用がかかるという問題があり、隨時様々な文章を読み上げさせるには、読み上げ対応能力の低いものとなっていた。

[0004] 一方、規則合成方式では、文脈や単語の差異を無視して、信号処理による音データを一文字ずつ対応させた繋ぎ合わせによって、それぞれ単音を順次連続されることで、肉声の音声波形に近い音を合成するようにしているために、出力される音は、機械的な音となってしまい、音質の低下は避けられず、その機械的な音は、自然な発声とは程遠いために聞いていて違和感の残るものとなっていた。

[0005] そこで、近年、人の肉声を文章として数多く収録したものが事前にデータベース化されて、これら膨大な音声データのデータベース(コーパス)を音源として音声を合成

する、例えば、以下に示す特許文献1及び2に開示されたコーパスベース音声合成技術が知られている。

[0006] 特許文献1:特許第2894447号公報

特許文献2:特許第2975586号公報

[0007] これら特許文献1及び2に開示されたコーパスベース音声合成技術では、データベース中に収録された多くの文章から必要な音素を抽出し、それら音素の組み合わせにより数多くの文章の合成が可能であることから、出力可能な合成文章数は膨大なものとなり、また、人間の肉声を音源に採用しているため、機械音による合成音声と比べて、より人間の肉声に近い自然な音声を出力することができる。

[0008] さらに、これら特許文献1あるいは特許文献2に開示されたコーパスベース音声合成技術によれば、新たに文章を追加して合成させる場合においても、既に収録されている音源中の音素を利用して合成が可能なために、データベースの追加録音も不要であり、したがって、追加コストはかかるないことから、現在、コールセンター等への導入が進められている。

### 発明の開示

#### 発明が解決しようとする課題

[0009] しかしながら、従来のコーパスベース音声合成技術では、任意の文章の合成に対応させるために、多数の音素を含む文章を収録するそのデータベースは膨大なものとなってしまい、装置の大型化が避けられず、例えば、コールセンター等に導入される場合には、業務内容対応用や、カタログ請求対応用、担当部署対応用等の、応対内容ごとのそれぞれ専用のデータベースを構築する必要があった。

[0010] また、装置が大型なものとなるために、小型製品の、例えば、聴覚障害者用の福祉関連機器や、玩具、家電等に内蔵することは困難なものであり、その用途は、例えば、コールセンター等に限定されてしまい、その導入は大規模な設備を備えた企業等のみに限定されたものとなっていた。

[0011] ここにおいて、本発明の解決すべき主要な目的は、次のとおりである。

[0012] 即ち、本発明の第1の目的は、コーパスベース音声合成を行う装置を小型化して、他の機器に組込み搭載可能な音声合成装置を提供せんとするものである。

[0013] 本発明の第2の目的は、ユーパスベース音声合成に採用される音声データベースを、用途ごとに選択収録された音声データを収録してそれぞれ着脱可能な音声合成装置を提供せんとするものである。

[0014] 本発明の他の目的は、明細書、図面、特に、特許請求の範囲の各請求項の記載から、自ずと明らかになろう。

### 課題を解決するための手段

[0015] 本発明装置においては、他の機器に組込み搭載される組込みマイコンからなる音声合成装置であつて、テキストデータ中の任意の文章を解析し、当該文章に対応する表音記号データを生成するテキスト解析部と、アクセントと抑揚について事前に設定された韻律知識ベースに従い、当該テキスト解析部で解析した文章の表音記号データのそれぞれに対応するアクセントと抑揚を示す韻律パラメータを生成する韻律予測部と、当該音声合成装置の用途に応じた所要の音声単位のみを有するよう事前に選択収録された所定の音声データのみを複数蓄積する音声データベースに基づき、当該韻律予測部にて生成した韻律パラメータのそれぞれに最も近い所定の音声単位部分を有した音声データのそれぞれから、該当する当該所定の音声単位部分の音声素片波形データの全てを抽出する音声単位抽出部と、当該音声単位抽出部にて抽出した音声素片波形データ群を、文章順に、当該音声素片波形データ群の示す音声波形が連続となるよう順次波形接続することで合成音声データを生成する波形接続部と、を具備させる、という特徴的構成手段を講じる。

[0016] さらに、具体的詳細に述べると、当該課題の解決では、本発明が次に列挙する上位概念から下位概念に亘る新規な特徴的構成手段を採用することにより、前記目的を達成するよう為される。

[0017] 即ち、本発明装置の第1の特徴は、事前に収録された所定文章の音声データを所定の音声単位毎に音声素片波形データとして抽出可能に複数蓄積した音声データベースを搭載し、任意のテキストデータに対して前記音声データベースに基づきユーパスベース音声合成を行うための音声合成装置であつて、シリアルデータからテキストデータを取得するデータ入力部と、当該テキストデータ中の前記任意の文章に対応する音を母音と子音で示す表音記号を表音記号データとして生成処理するテキス

ト解析部と、アクセントと抑揚について事前に設定された韻律知識ベースに従い、事前に解析された前記テキストデータ中の任意の文章に対応する前記表音記号データのそれぞれに対応する前記アクセントと前記抑揚を示す韻律パラメータを生成する韻律予測部と、当該音声合成装置の用途に応じた所要の前記音声単位のみを有するよう事前に選択収録された所定の前記音声データのみを複数蓄積する前記音声データベースに基づき、当該韻律予測部にて生成した前記韻律パラメータのそれぞれに最も近い前記所定の音声単位部分を有した前記音声データのそれぞれから、該当する当該所定の音声単位部分の前記音声素片波形データの全てを抽出する音声単位抽出部と、当該音声単位抽出部にて抽出した前記音声素片波形データ群を、前記文章順に、当該音声素片波形データ群の示す音声波形が連続となるよう順次波形接続することで合成音声データを生成する波形接続部と、当該合成音声データをアナログ音声に変換出力する音声変換処理部と、を具備してなる、音声合成装置の構成採用にある。

- [0018] 本発明装置の第2の特徴は、上記本発明装置の第1の特徴における前記音声データベースが、前記音声合成装置に着脱可能なメモリカード上に構築されて、当該音声合成装置に当該メモリカードが装着されたときに、前記音声単位抽出部から読み取り可能に構成されてなる、音声合成装置の構成採用にある。
- [0019] 本発明装置の第3の特徴は、上記本発明装置の第1の特徴における前記データ入力部が、当該音声合成装置が組込み搭載された他の機器と接続されて、当該機器からシリアルデータを受信してなる、音声合成装置の構成採用にある。
- [0020] 本発明装置の第4の特徴は、上記本発明装置の第1の特徴における前記音声合成装置が、前記データ入力部により前記任意文章とともに取得した速度パラメータを前記波形接続部より生成された前記合成音声データに反映させ、当該合成音声データの読み上げ速度を調整する話速変換部を前記音声変換処理部に前置してなる、音声合成装置の構成採用にある。
- [0021] 本発明装置の第5の特徴は、上記本発明装置の第1の特徴における前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部が、一個のケース内に一体装備

されてなる、音声合成装置の構成採用にある。

[0022] 本発明装置の第6の特徴は、上記本発明装置の第1の特徴における前記データ入力部と前記波形接続部と前記音声変換処理部が、他の機器に組込み搭載される組込みマイコンに一体装備し、前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベースと前記音声単位抽出部が、センター内パソコン 컴퓨터に装備され、前記組込みマイコンと当該センター内パソコン 컴퓨터は、同一ネットワーク上に別途設置され、当該センター内パソコン 컴퓨터内で前記データ入力部、前記テキスト解析部、前記韻律予測部、前記音声データベースに直結した前記音声単位抽出部を経て前記テキストデータから変換された前記音声素片波形データを、前記ネットワークを介して前記組込みマイコンの前記波形接続部に送信可能とし、当該組込みマイコンの前記音声変換処理部へ当該波形接続部から合成音声を受渡しするシステムに構築されてなる、音声合成装置の構成採用にある。

[0023] 本発明装置の第7の特徴は、上記本発明装置の第1の特徴における前記音声合成装置が、別途配置された任意のパソコン 컴퓨터に前記データ入力部が接続されて、前記テキスト解析部にて解析する前記テキストデータを当該パソコン 컴퓨터から取得可能に構成される一方、前記音声変換処理部として別途配置された任意のスピーカに接続されて、前記波形接続部にて生成した前記合成音声データを当該スピーカにて音声出力可能に構成されてなる、音声合成装置の構成採用にある。

[0024] 本発明装置の第8の特徴は、上記本発明装置の第1の特徴における前記所定の音声単位が、音素、単語、フレーズ、シラブルのうちの1以上である、音声合成装置の構成採用にある。

[0025] 本発明装置の第9の特徴は、上記本発明装置の第1の特徴における前記データ入力部と前記テキスト解析部が、初期設定時のみ使用のパソコン 컴퓨터に装備してシリアルデータを入力し表音記号データを出力する初期設定機能をそれぞれ有し、前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部が、その他の機器に組込み搭載される組込みマイコン

ンに装備し、前記パーソナルコンピュータを当該組込みマイコンに初期設定時のみ接続して、当該パーソナルコンピュータから出力した前記表音記号データを当該組込みマイコンの前記韻律予測部に入力するとともに、前記音声データベースに予め設定記録しておき、当該組込みマイコンに入力したシリアルデータを当該韻律予測部と当該音声データベースに直結する前記音声単位抽出部と前記波形接続部と前記音声変換処理部を順次経てアナログ出力する機能を有してなる、音声合成装置の構成採用にある。

[0026] 本発明装置の第10の特徴は、上記本発明装置の第1の特徴における前記データ入力部と前記波形接続部と前記音声変換処理部が、緊急警報用又は、案内、連絡用出力端末に組込みマイコンとして組込まれるとともに、前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベースと前記音声単位抽出部は、これら等を内蔵装備したセンター内パーソナルコンピュータとして、前記組込みマイコンとネットワークを介し一方向送信可能にシステム構築してなる、音声合成装置の構成採用にある。

[0027] 本発明装置の第11の特徴は、上記本発明装置の第1の特徴における前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部が、前記データ入力部及び前記テキスト解析部と初期設定後切り離して玩具やその他の機器に組込みマイコンとして組み込まれてなる、音声合成装置の構成採用にある。

## 発明の効果

[0028] 本発明によれば、従来、大型化が避けられなかったユーパスベース音声技術を採用した音声合成装置を組込みマイコンにて構成し、従来と比較して大幅に小型化することが可能となり、他の機器に組込み可能となることから、例えば、福祉関連機器に組み込まれることで音声伝達を可能とするコミュニケーションツールとして活用することができ、キャラクターの声を出力可能な人形等の玩具や、音声による情報伝達が可能な家電等、さまざまな製品に活用することが可能となる。

[0029] また、音声データベースを着脱可能なメモリカード上に構築し用途に応じて交換可能にしたことで、音声合成装置の小型化が可能となると共に、用途に合った音声デ

ータを収録することで、音声合成の読み正解率やアクセント正解率の向上させてより自然な音声を出力させることが可能になり、また、出力する声質をユーザの好みのものに切り替えることが可能となる。

[0030] さらに、ネットワークを利用して音声合成を行う場合、従来音声の送信には中高速回線を利用していたが、本発明では、テキストデータを受信側装置で受信して音声に変換すればよいため、低速回線を利用した音声放送が可能となり、また、プッシュ型のサービスに適用された場合、テキストデータのみを配信することで受信側装置にて音声として出力させることができ、省力化や、防災無線等の緊急を要する場合にも迅速なサービスの提供が可能となる。

#### 図面の簡単な説明

[0031] [図1]本発明の一形態例に係る音声合成装置の機能構成図である。  
 [図2]同上した音声合成装置に話速変換部を機能追加した音声合成装置の機能構成図である。  
 [図3]同上した音声合成装置のハードウェア構成例を示す模式図である。  
 [図4]同上した音声合成装置のデータ構成を説明するための図であり、同図(a)はテキストデータを、同図(b)は表音記号データを、同図(c)は韻律知識ベースを、同図(d)は韻律パラメータを、同図(e)は音声データベースを説明するための図である。  
 [図5]本発明の機能構成例2に係る音声合成装置の機能構成図である。  
 [図6]本発明の機能構成例3に係る音声合成装置の機能構成図である。  
 [図7]本発明の形態例に係る音声合成装置をパソコン用コンピュータに搭載させたハードウェア構成例を示す模式図である。

#### 符号の説明

[0032]  $\alpha$ 、 $\alpha 1$ …音声合成装置  
 $\alpha 2$ 、 $\alpha 3$ …組込みマイコン  
 $\beta$ 、 $\beta 2$ 、 $\beta 3$ …パソコン用コンピュータ  
 $\gamma$ …音声合成システム  
 1…音声データベース  
 2…テキスト解析部

3…韻律予測部

3A…韻律知識ベース

4…音声単位抽出部

5…波形接続部

6…データ入力部

7…音声変換処理部

8…話速変換部

11…CPU

12…ROM

13…RAM

14…メモリカード

15…シリアルインターフェース

16…D/Aコンバータ

21…入力手段

22…スピーカ

### 発明を実施するための最良の形態

[0033] 以下、本発明の実施の形態につき、添付図面を参照しつつ、音声合成装置の形態例を説明する。

[0034] (形態例)

まず、図1は、本発明の一形態例に係る音声合成装置の機能構成図である。

[0035] 同図に示すように、本形態例にかかる音声合成装置 $\alpha$ は、事前に収録された所定文章の音声データを、例えば、音素や単語、フレーズ、シラブル等の所定の音声単位毎に音声素片波形データとして抽出可能に複数蓄積した音声データベース1を搭載し、任意のテキストデータに対して音声データベース1に基づきコーパスベース音声合成を行うための装置であり、少なくともテキスト解析部2と、韻律予測部3と、音声単位抽出部4と、波形接続部5からなり、必要に応じ他の機器に組込み搭載される組込みマイコンとして構成する。

[0036] なお、上記機能部全てをマイコンが具備するよう限定される必要はなく、使用用途

及びその規模に応じて複数の所定機能部をマイコンに具備させ、その他の各機能部をパーソナルコンピュータに実行させるなどして構成することもできる。

- [0037] ここで、音声データベース1は、コーパスベース音声合成を行うためのコーパスであって、音声合成装置 $\alpha$ の用途に応じた所定の音声単位のみを有するように事前に選択収録された所定の音声データのみを複数蓄積し、音声合成装置 $\alpha$ の用途に応じて細分化されて構築される。
- [0038] 一方、テキスト解析部2は、入力されたテキストデータ中の任意の文章を解析し、この文章に対応する表音記号データを生成可能に構成されて、韻律予測部3は内部に、表音記号データのアクセントと抑揚についての認識規則について事前に設定された韻律知識ベース3Aを搭載し、この韻律知識ベース3Aに従い、テキスト解析部2が生成した表音記号データのそれぞれに対応するアクセントと抑揚を示す韻律パラメータを生成するよう構成される。
- [0039] 他方、音声単位抽出部4は、韻律予測部3にて生成された韻律パラメータのそれぞれに最も近いアクセントと抑揚を備えた音素を含む音声データを、例えば、人間の聴覚特性に近づけた評価関数等を用いて、音声データベース1から抽出し、ここで抽出した音声データのそれから、この韻律パラメータに該当する音素等の所定の音声単位の音声素片波形データのみを抽出するよう構成される。
- [0040] さらに、波形接続部5は、音声単位抽出部4にて抽出した複数の音声素片波形データ群を、文章順に、この音声素片波形データ群の音声波形が滑らかで自然な音声となるよう、連続させて順次波形接続を行うことで、自然な韻律の合成音声データを生成するよう構成される。
- [0041] なお、音声合成装置 $\alpha$ の組込みマイコン上に、さらに、音声合成装置 $\alpha$ が組込み搭載される他の機器と接続されて、この機器中の、例えば、キーボードやマウス等の入力手段や、ネットワークを介して送受信されたデータを記録する記録媒体等からシリアルデータを受信し、このシリアルデータからテキストデータを取得してテキスト解析部2に入力可能に構成されたデータ入力部6を具備するようにしても構わない。
- [0042] このデータ入力部6を具備することにより、音声合成装置 $\alpha$ は、事前に設定されたテキストデータの音声合成のみならず、例えば、音声合成装置 $\alpha$ のユーザにより入

力された任意の文章の音声合成が可能となり、ユーザからの任意のテキストデータの入力に対応可能となるとともに、所望の文章を随時受付けて即座に合成音声として出力する等のリアルタイム性を確保することが可能となる。

- [0043] また、音声合成装置 $\alpha$ の組込みマイコン上に、波形接続部5にて生成した合成音声データをアナログ変換し、このアナログ変換した合成音声データを別途接続されたスピーカ等にアナログ出力することで、合成音声データを音声出力する音声変換処理部7を具備するようにしても構わない。
- [0044] なお、音声合成装置 $\alpha$ が組み込まれる他の機器に、データ入力部6と音声変換処理部7とに代わる同様の機能を備えたインターフェースやコンバータ等が搭載された場合、音声合成装置 $\alpha$ は、音声合成装置 $\alpha$ 内にデータ入力部6と音声変換処理部7とを搭載させずに、テキストデータを取得可能とともに、合成音声データを音声出力することが可能となるようにしても構わない。
- [0045] さらに、図2は図1の音声合成装置 $\alpha$ に合成音声の読み上げ速度の調節機能を加えた構成図である。
- [0046] 同図に示すように、音声合成装置 $\alpha$ 1が組込み搭載される他の機器からテキストデータとともに入力された速度パラメータを、波形接続部5により生成された合成音声データに反映させ、合成音声の読み上げ速度を調節する話速変換部8を音声合成装置 $\alpha$ 1のマイコン上に具備してもよい。
- [0047] 次に、図3は、本形態例に示した音声合成装置 $\alpha$ のハードウェア構成例を示す模式図である。
- [0048] 同図に示すように、音声合成装置 $\alpha$ は、音声合成装置 $\alpha$ における各機能部をそれぞれ一連に制御するCPU(Central Processing Unit)11と、CPU11からアクセス可能なROM(Read Only Memory)12と、RAM(Random Access Memory)13とを搭載して構成されて、例えば、ROM12にリアルタイムOS(Operating System)と、テキスト解析部2、韻律予測部3、音声単位抽出部4、波形接続部5の各機能を音声合成装置 $\alpha$ のCPU11に実行させるための処理プログラム等が記録されるとよい。
- [0049] さらに、音声合成装置 $\alpha$ は、例えば、フラッシュメモリ等で構成されて音声合成装置

$\alpha$ に着脱可能なメモリカード14を着脱可能に構成されて、このメモリカード14に音声データベース1を構築することにより、音声合成装置 $\alpha$ が組み込まれる機器の用途や、音声合成装置 $\alpha$ を利用するユーザの好みにより、所要のメモリカード14への差換えが可能になり、装着されたメモリカード14内の音声データベース1に基づき音声単位抽出部4が機能するように構成されるとよい。

- [0050] また、データ入力部6として機能するシリアルインターフェース15や、音声変換処理部7として機能するD/Aコンバータ16(D/A:Digital to Analog)を搭載するようにしても構わない。
- [0051] 続いて、図4は、本形態例に示した音声合成装置 $\alpha$ のデータ構成を説明するための図であり、同図(a)はテキストデータを、同図(b)は表音記号データを、同図(c)は韻律知識ベースを、同図(d)は韻律パラメータを、同図(e)は音声データベースを説明するための図であり、アクセントや抑揚については説明のために模式的に示している。
- [0052] 同図(a)に示すように、テキスト解析部2に入力されるテキストデータは、例えば、データ入力部6にて取得したシリアルデータ中の「橋を渡る」のような任意の文章であり、ここで、このテキストデータは、かな・漢字等の混合であっても構わず、音声にすることが可能なものであれば、そのテキストデータに使用される文字は限定されるものではない。
- [0053] なお、テキストデータは、テキスト形式のデータファイルに限らず、HTML(Hyper Text Markup Language)形式のデータファイルからHTMLタグを取り除いて抽出されたものでも構わず、インターネット上のホームページや、電子メール、あるいは、キーボードやマウス等の入力手段によってユーザから直接入力されて生成されたテキストデータであっても構わない。
- [0054] 一方、同図(b)に示すように、テキスト解析部2で生成する表音記号データは、例えば、テキストデータの音を母音と子音で示すような表音記号を採用して、同図(a)にて示したテキストデータに基づき生成される表音記号データは、例えば、「ha shi wo wa ta ru」となる。
- [0055] また、韻律知識ベース3Aは、表音記号データのアクセントや抑揚等を決定するた

めに韻律予測部3にて利用される予め設定された規則であって、例えば、同図(b)に示した表音記号データの「ha shi」について、日本語の「橋」、「端」、「箸」、…のいずれに対応するものであるかを文脈から決定し、これら表音記号データのアクセントや抑揚について決定可能なアルゴリズムを備える。

[0056] そこで、韻律予測部3は、例えば、韻律知識ベース3Aに基づき「橋」に対応する表音記号データ中の「ha shi」について所定の音声単位毎の、ここでは、「ha」と「shi」に対してそれぞれ韻律パラメータを生成可能に構成されて、韻律知識ベース3Aに従い、表音記号データの全てについて、アクセントや抑揚、さらには、音声と音声との間、音声のリズム、スピード等を決定することができる。

[0057] なお、ここでは、アクセントと抑揚を説明のために模式的に、表音記号に重ねて下線や上線等で示して説明するものの、音声合成装置 $\alpha$ 内にてアクセントや抑揚等の音声に必要な情報を識別可能に記録されれば、どのような形態であっても構わない。

[0058] 他方、同図(d)に示すように、韻律予測部3にて同図(c)で説明した韻律知識ベース3Aに従い生成される韻律パラメータは、各表音記号に対応させてテキストデータの文脈に合った、例えば、アクセント、抑揚及び音声間の間についてそれぞれパラメータとして示すものであり、図に示した「wo」と「wa」のアクセントを示す下線間の切れ目は、その表音記号間に所定間隔の間を示すものである。

[0059] 続いて、同図(e)に示すように、音声単位抽出部4からアクセスされる音声データベース1は、予め複数の所定文章の肉声がそのアクセントと抑揚等の韻律知識ベース3Aに対応させた音声データとともに、例えば、音素等の所定の音声単位毎に音声素片波形データとして抽出可能に蓄積されて、ここでは、「春が来た」、「使用する」、「映画を見る」、「私は」、…等の音声データが蓄積されていることを示している。

[0060] したがって、音声単位抽出部4は、韻律予測部3から同図(d)に示すような韻律パラメータを受信したときには、この韻律パラメータが示すそれぞれ固有のアクセントと抑揚を有した「ha」、「shi」、「wo」、「wa」、「ta」、「ru」のそれぞれと、表音記号が対応するとともに最も近いアクセントと抑揚を有した音声データを音声データベース1から検索する。

[0061] 引続き、音声単位抽出部4は、ここで抽出した「春が来た」、「使用する」、「映画を見

る」、「私は」、…等の音声データから、韻律パラメータに合致する「ha」、「shi」、「wo」、「wa」、「ta」、「ru」の音声素片波形データのみを切り取り、抽出するようことで、波形接続部5にてこの音声素片波形データを滑らかに接続して、合成音声データを生成することが可能となる。

[0062] なお、ここでは、所定の音声単位の一例として音素を採用した場合について説明したが、入力されたテキストデータ中に事前に音声データベース1中に蓄積された単語やフレーズを含む場合には、所定の音声単位をこの単語やフレーズとすることで、音声単位抽出部4において、音声データベース1中に蓄積された単語やフレーズ等を分割せずにそのまま抽出することが可能であり、この単語あるいはフレーズをそのまま、あるいは、組み合わせて出力することで、より自然な音声を合成することが可能となる。

### 実施例

[0063] 続いて、実施例として図1、2の機能構成図及び図5、6に示す本発明における合成音声装置 $\alpha$ の実施構成図を用いて機能構成例を説明する。

[0064] (構成例1)

まず、構成例1としては上記形態例にて説明した、図1の機能構成図に示す各機能部1～7全てをマイコンに組込んだ音声合成装置 $\alpha$ を機能構成例1とする。

[0065] この場合、全機能部1～7は一個のケース内に一体装備され、他の設備、機器等に機能を分散しなくても単体のみで音声の合成が実行可能な音声合成装置 $\alpha$ であり、シリアルデータ入力からアナログ出力までの一連の機能部1～7を一個体のケースで実行させることを可能とする。

[0066] また、上記全機能部がケース単体で実行可能であればその機能構成は限定されず、例えば、同一ケース内に音声変換出力部7、データ入力部6として図示しないスピーカやデータ入力装置等を組込み搭載してもよい。

[0067] (構成例2)

さらに、構成例1の音声合成装置 $\alpha$ に合成音声の読み上げ速度調節機能である話速変換部8を付加し、構成例1と同様に図2に記載される全機能部1～8を一個のケースに一体装備した音声合成装置 $\alpha$ 2を機能構成例2とする。

[0068] ここで話速変換部8は、合成音声データに速度パラメータを反映させることにより合成音声の速度調節を行う。この場合、データ入力部にはシリアルデータとして、テキストデータとともに速度パラメータを入力する。

[0069] 速度パラメータはデータ入力部6から波形接続部5までの機能部を各変換データ及びパラメータに付加された状態で受け渡され、話速変換部8ではじめて認識される。話速変換部8は、波形接続部5から速度パラメータとともに受け取った合成音声データに対し速度パラメータの値をあてはめ、合成音声の読み上げ速度の変更を行う。

[0070] 構成例2は話速変換を行うことにより、利用状況に応じて速度の変更を行い利用者に対し正確に合成音声を伝達することを目的としており、例えば、読み上げ速度を通常よりも遅く設定することにより、聞き取り易くして緊急時など冷静な判断能力が欠如しがちな状態において効果的である。

[0071] (構成例3)

次に図5は、図1に示す音声合成装置 $\alpha$ における波形接続部5、音声変換処理部7を抽出選択し組込みマイコン $\alpha$ 2に搭載し、他の各機能部を別途設置したパーソナルコンピュータに組み込むことにより一連の音声合成処理を実行する音声合成システム $\gamma$ の構成例を示す機能構成図である。

[0072] 同図に示すように、本構成例3にかかる音声合成システム $\gamma$ は、組込みマイコン $\alpha$ 2として火災、地震などの災害時等に入力されたテキストデータを合成音声に変換し緊急警報として使用する場合に用いられる出力端末を目的とする音声合成システムの一例である。

[0073] 図5に示すように、音声合成システム $\gamma$ は、波形接続部5、音声変換処理部7を具備する組込みマイコン $\alpha$ 2と、図1に図示する上記以外の各機能部である音声データベース1とデータ入力部6から音声単位抽出部4までの各機能部を具備したパーソナルコンピュータ等のマシンとをネットワーク接続して使用する。

[0074] 組込みマイコン $\alpha$ 2は単体としてネットワークに接続するか、又は、他の機器に組込みマイコン $\alpha$ 2を組込み搭載して利用するとよい。

[0075] 上記ネットワークの接続方法は、一般的に家庭内や小規模な設備などでも容易に接続可能なインターネット回線や電話回線などが候補として挙げられるが、他に無線

、専用回線など、別途設置してある設備とのデータの通信が可能なものであれば、その接続手段は限定されない。

[0076] 図1に示す音声合成装置 $\alpha$ に記載される各機能部の中で、負荷が高くデータ処理時間を要するデータ入力部6から音声単位抽出部4までの機能を、別途設置する高速処理可能な大容量のパーソナルコンピュータ $\beta$ 2のマシンに実行させ、ネットワークを介して組込みマイコン $\alpha$ 2がパーソナルコンピュータ $\beta$ 2から受信した音声素片波形データを合成音声データへの変換処理のみを実行することにより、緊急を要する状態においても高速な音声合成処理が可能となるという優れた効果を奏する。

[0077] 本構成例2は緊急警報用ばかりでなく案内、連絡用として利用してもよく、また構成例2で示す話速変換部8を本構成例に具備させることにより、状況に応じて読み上げ速度の変更を行うことも可能である。

[0078] (構成例4)

次に、図6は、図5と同様に図1に示す音声合成装置 $\alpha$ の機能部1と3～5、7を一部抽出選択した組込みマイコン $\alpha$ 3の機能構成図である。

[0079] 同図に示すように、本構成例4にかかる組込みマイコン $\alpha$ 3は、データ入力部6及びテキスト解析部2を組み込んだ任意のパーソナルコンピュータ $\beta$ 3にから表音記号データを取得可能な構成にし、音声データベース1と、合成音声として出力する韻律予測部3から音声変換処理部7までの一連の機能部とを組み込んだマイコンである。なお、パーソナルコンピュータ $\beta$ 3は初期設定後は切り離される。

[0080] 組込みマイコン $\alpha$ 3は、玩具のような小型の装置等その他の機器に搭載されることを目的とし、搭載する装置の例としては、玩具、携帯電話、補聴器等の福祉関連機器その他が挙げられる。

[0081] これらの装置は、利用者に対し合成音声の提供を実現するものであるが、入力されるシリアルデータの内容はある程度定まっているため、予めテキスト解析を行うことにより処理効率の向上を図ることができる。

[0082] また、これらは上述したような小型装置のみに限定されず、例えば自動販売機や、カーナビゲーションシステム、無人受付設備など、出力される合成音声の内容が限定されている装置に利用すれば、新たに大きな設備を設置することなく、組込みマイ

コン $\alpha$ 3を追加搭載させるだけで合成音声機能をこれらの装置に組み込むことが可能となる。

- [0083] 続いて、図7は、本形態例に示した音声合成装置 $\alpha$ を、他の機器としてのパソコン用コンピュータ $\beta$ に搭載させたハードウェア構成例を示す模式図である。
- [0084] 同図に示すように、音声合成装置 $\alpha$ は、別途配置された任意のパソコン用コンピュータ $\beta$ に搭載されて接続された場合、例えば、パソコン用コンピュータ $\beta$ に搭載された入力手段21からデータ入力部6にてシリアルデータを受信可能に構成される一方、音声合成装置 $\alpha$ にてシリアルデータに基づき生成した合成音声データを、音声変換処理部7から別途パソコン用コンピュータ $\beta$ に内蔵された音声出力可能なスピーカ22にアナログ出力することで、このスピーカ22にて音声出力させることが可能となる。
- [0085] このとき、音声合成装置 $\alpha$ は、この音声合成装置 $\alpha$ 内に、事前に音声データベース1を記録するメモリカード14が装着されるとよく、メモリカード14は、音声合成装置 $\alpha$ に事前に固定的かつ専用に搭載されたものであっても、パソコン用コンピュータ $\beta$ を利用するユーザにより任意に他のメモリカード14と差換え可能とするようにしても構わない。
- [0086] 以上、本発明の実施の形態につき、その音声合成装置 $\alpha$ の一形態例及び機能構成例を挙げて説明したが、本発明は、必ずしも上述した手段にのみ限定されるものではなく、前述した効果を有する範囲内において、適宜、変更実施することが可能なものである。
- [0087] また、音声合成装置 $\alpha$ を、別途構成された他の音声認識装置と接続することにより、自然な発声の対話を可能とする対話型の音声合成装置を構築することも可能である。

## 請求の範囲

[1] 事前に収録された所定文章の音声データを所定の音声単位毎に音声素片波形データとして抽出可能に複数蓄積した音声データベースを搭載し、任意のテキストデータに対して前記音声データベースに基づきコーパスベース音声合成を行うための音声合成装置であつて、

シリアルデータからテキストデータを取得するデータ入力部と、

当該テキストデータ中の前記任意の文章に対応する音を母音と子音で示す表音記号を表音記号データとして生成処理するテキスト解析部と、

アクセントと抑揚について事前に設定された韻律知識ベースに従い、事前に解析された前記テキストデータ中の任意の文章に対応する前記表音記号データのそれぞれに対応する前記アクセントと前記抑揚を示す韻律パラメータを生成する韻律予測部と、

当該音声合成装置の用途に応じた所要の前記音声単位のみを有するよう事前に選択収録された所定の前記音声データのみを複数蓄積する前記音声データベースに基づき、当該韻律予測部にて生成した前記韻律パラメータのそれぞれに最も近い前記所定の音声単位部分を有した前記音声データのそれから、該当する当該所定の音声単位部分の前記音声素片波形データの全てを抽出する音声単位抽出部と、

当該音声単位抽出部にて抽出した前記音声素片波形データ群を、前記文章順に、当該音声素片波形データ群の示す音声波形が連続となるよう順次波形接続することで合成音声データを生成する波形接続部と、

当該合成音声データをアナログ音声に変換出力する音声変換処理部と、を具備する、

ことを特徴とする音声合成装置。

[2] 前記音声データベースは、

前記音声合成装置に着脱可能なメモリカード上に構築されて、

当該音声合成装置に当該メモリカードが装着されたときに、前記音声単位抽出部から読み取り可能に構成される、

ことを特徴とする請求項1に記載の音声合成装置。

[3] 前記データ入力部は、

当該音声合成装置が組込み搭載された他の機器と接続されて、当該機器からシリアルデータを受信する、

ことを特徴とする請求項1に記載の音声合成装置。

[4] 前記音声合成装置は、

前記データ入力部により前記任意文章とともに取得した速度パラメータを前記波形接続部より生成された前記合成音声データに反映させ、当該合成音声データの読み上げ速度を調整する話速変換部を前記音声変換処理部に前置する、

ことを特徴とする請求項1に記載の音声合成装置。

[5] 前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部は、

一個のケース内に一体装備される、

ことを特徴とする請求項1に記載の音声合成装置。

[6] 前記波形接続部と前記音声変換処理部は、

他の機器に組込み搭載される組込みマイコンに一体装備し、

前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベースと前記音声単位抽出部は、

センター内パーソナルコンピュータに装備され、

前記組込みマイコンと当該センター内パーソナルコンピュータは、

同一ネットワーク上に別途設置され、

当該センター内パーソナルコンピューター内で前記データ入力部、前記テキスト解析部、前記韻律予測部、前記音声データベースに直結した前記音声単位抽出部を経て前記テキストデータから変換された前記音声素片波形データを、前記ネットワークを介して前記組込みマイコンの前記波形接続部に送信可能とし、当該組込みマイコンの前記音声変換処理部へ当該波形接続部から合成音声を受渡しするシステムに構築される、

ことを特徴とする請求項1に記載の音声合成装置。

[7] 前記音声合成装置は、  
別途配置された任意のパーソナルコンピュータに前記データ入力部が接続されて、前記テキスト解析部にて解析する前記テキストデータを当該パーソナルコンピュータから取得可能に構成される一方、  
前記音声変換処理部として別途配置された任意のスピーカに接続されて、前記波形接続部にて生成した前記合成音声データを当該スピーカにて音声出力可能に構成される、  
ことを特徴とする請求項1に記載の音声合成装置。

[8] 前記所定の音声単位は、  
音素、単語、フレーズ、シラブルのうちの1以上である、  
ことを特徴とする請求項1に記載の音声合成装置。

[9] 前記データ入力部と前記テキスト解析部は、  
初期設定時のみ使用のパーソナルコンピュータに装備してシリアルデータを入力し表音記号データを出力する初期設定機能をそれぞれ有し、  
前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部は、  
その他の機器に組込み搭載される組込みマイコンに装備し、  
前記パーソナルコンピュータを当該組込みマイコンに初期設定時のみ接続して、当該パーソナルコンピュータから出力した前記表音記号データを当該組込みマイコンの前記韻律予測部に入力するとともに、前記音声データベースに予め設定記録しておき、当該組込みマイコンに入力したシリアルデータを当該韻律予測部と当該音声データベースに直結する前記音声単位抽出部と前記波形接続部と前記音声変換処理部を順次経てアナログ出力する機能を有する、  
ことを特徴とする請求項1に記載の音声合成装置。

[10] 前記波形接続部と前記音声変換処理部は、  
緊急警報用又は、案内、連絡用出力端末に組込みマイコンとして組込まれるとともに、  
前記データ入力部と前記テキスト解析部と前記韻律予測部と前記音声データベー

スと前記音声単位抽出部は、

これら等を内蔵装備したセンター内パーソナルコンピュータとして、前記組込みマイコンとネットワークを介し一方向送信可能にシステム構築してなる、  
ことを特徴とする請求項1に記載の音声合成装置。

[11] 前記韻律予測部と前記音声データベースと前記音声単位抽出部と前記波形接続部と前記音声変換処理部は、

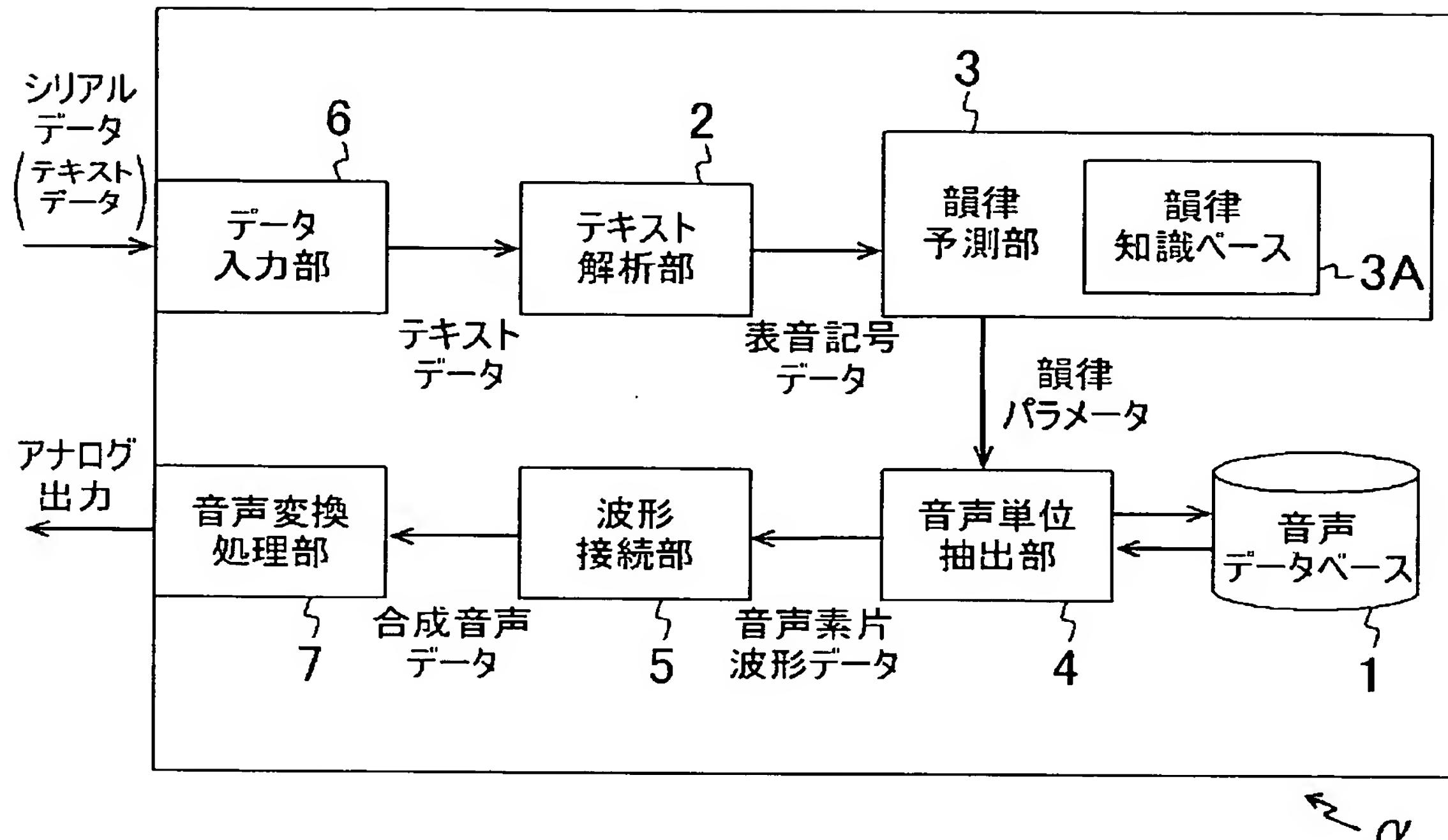
前記データ入力部及び前記テキスト解析部と初期設定後切り離して玩具やその他の機器に組込みマイコンとして組み込まれてなる、  
ことを特徴とする請求項1に記載の音声合成装置。

## 要 約 書

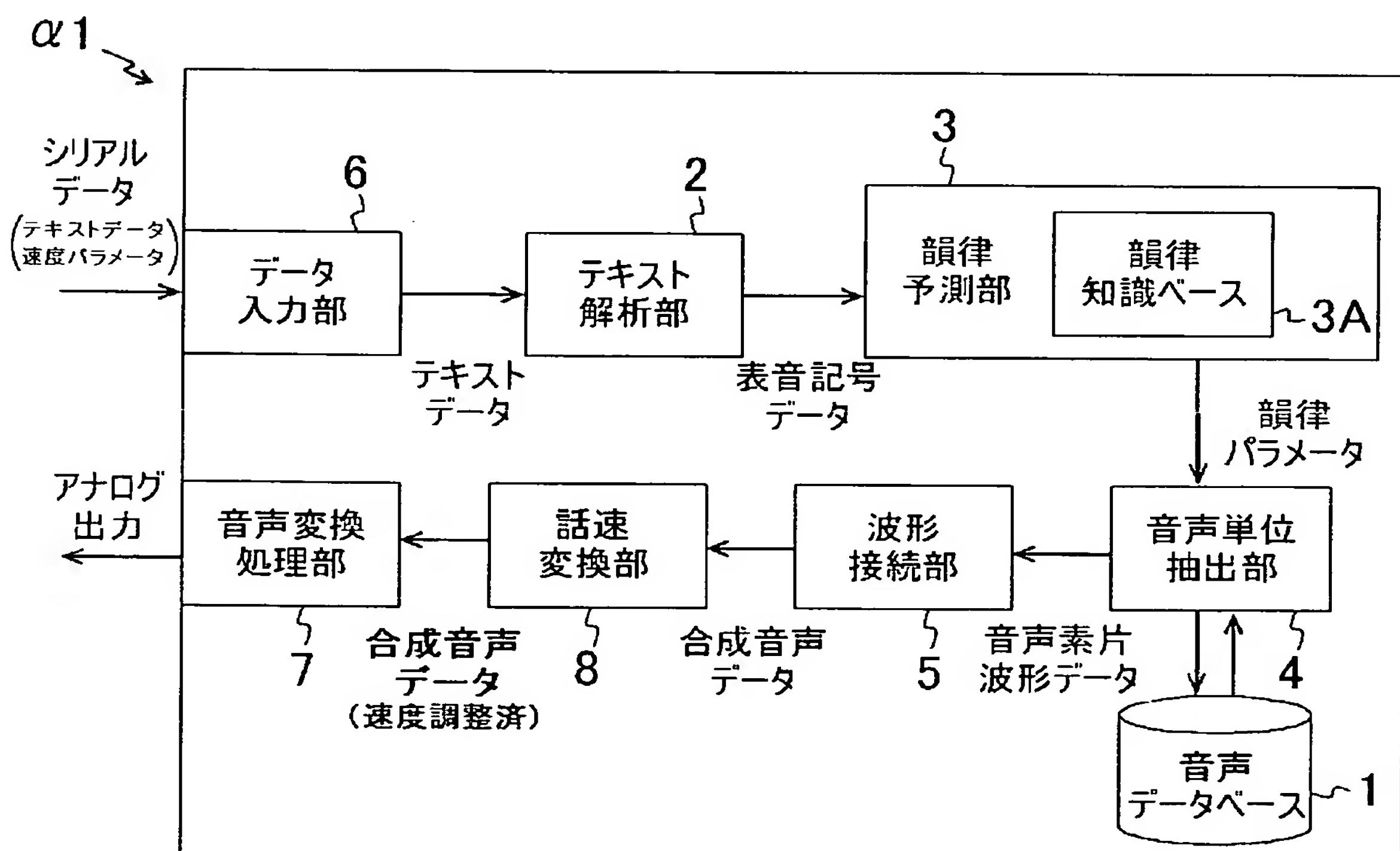
テキストデータ中の任意の文章を解析し対応する表音記号データを生成するテキスト解析部と、アクセントと抑揚について事前に設定された韻律知識ベースに従い、表音記号データのそれぞれに対応するアクセントと抑揚を示す韻律パラメータを生成する韻律予測部と、音声合成装置の用途に応じた所要の音声単位のみを有するよう事前に選択収録された所定の音声データのみを複数蓄積する音声データベースに基づき、韻律パラメータのそれぞれに最も近い音声単位部分を有した音声データのそれぞれから、所定の音声単位部分の音声素片波形データの全てを抽出する音声単位抽出部と、これら音声素片波形データ群を音声波形が連続となるよう順次波形接続することで合成音声データを生成する波形接続部と、データ入力部、音声変換処理部及び話速変換部も含め前記各機能部を使用用途及び装置の規模に応じて適宜具備又は切離すコーパスベース音声合成装置。

THIS PAGE BLANK (USPTO)

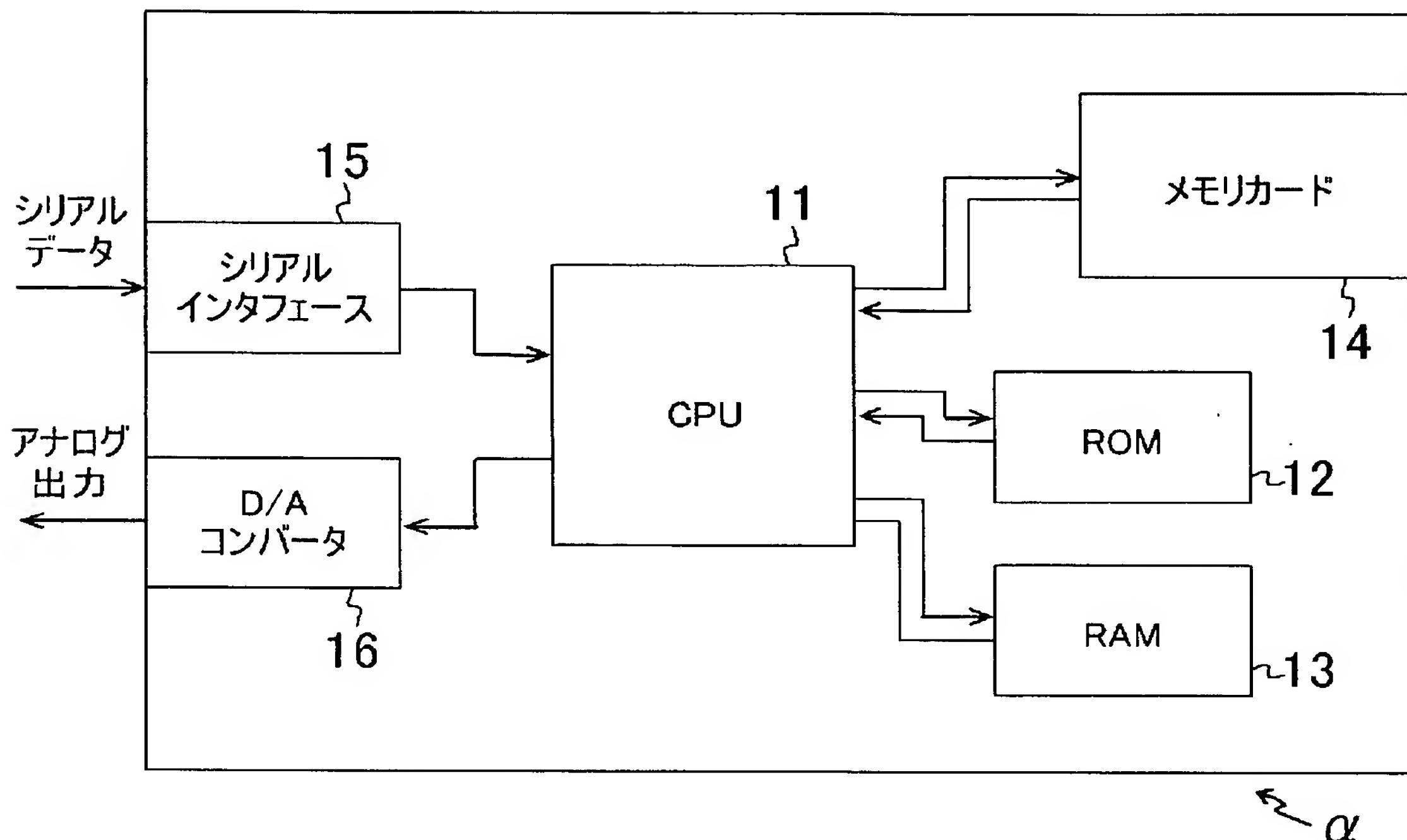
[図1]

 $\alpha$ 

[図2]

 $\alpha 1$

[図3]



[図4]

テキストデータ  
橋を渡る

(a)

韻律パラメータ  
ha shi wo wa ta ru

(d)

表音記号データ  
ha shi wo wa ta ru

(b)

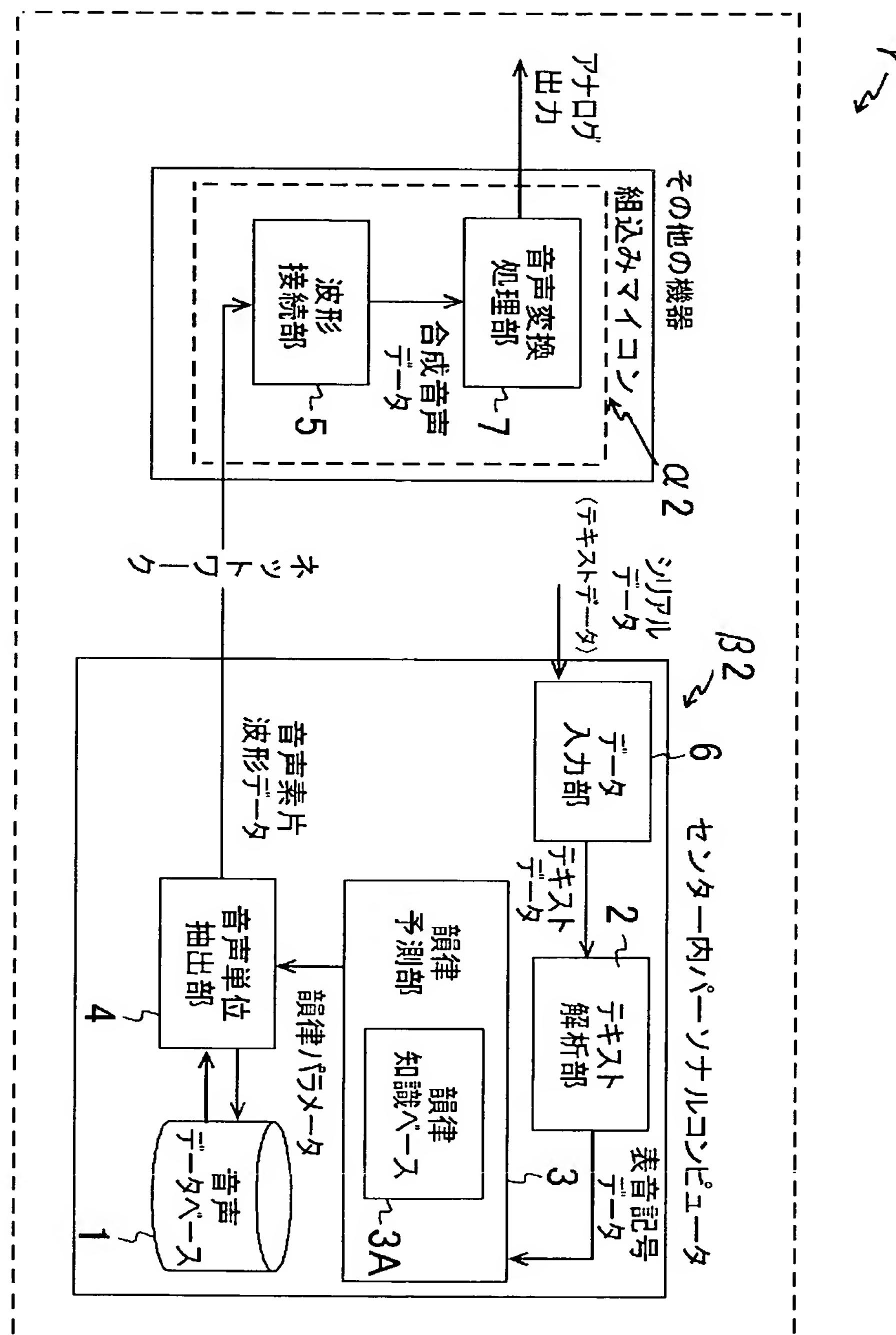
音声データベース  
春が来た: ha ru ...  
使用する: shi yo u ...  
映画を見る: ei ga wo ...  
私は: wa ta shi wa  
⋮

(e)

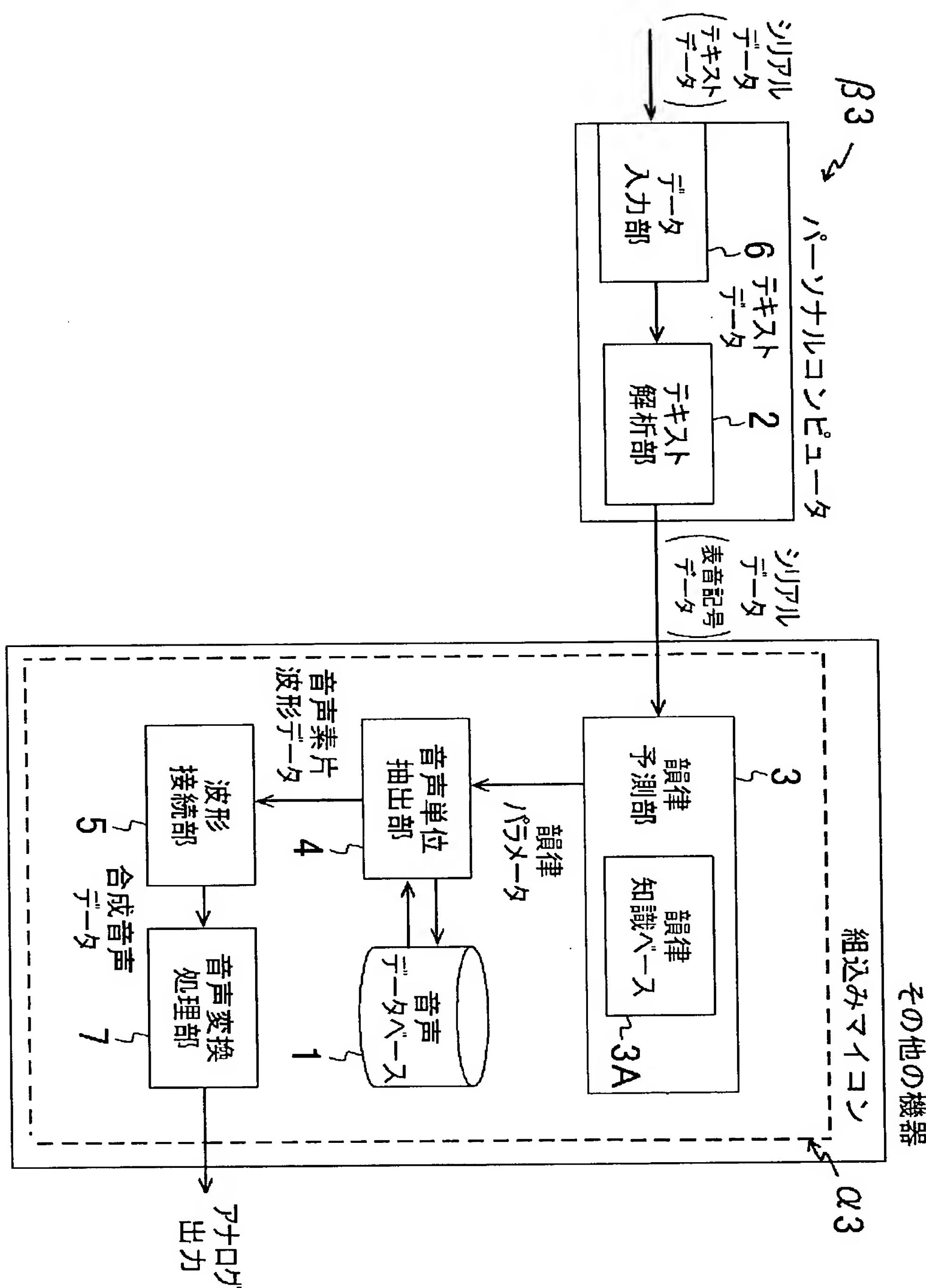
韻律知識ベース  
ha shi = 橋, 端, 箸, ...  
橋: ha shi

(c)

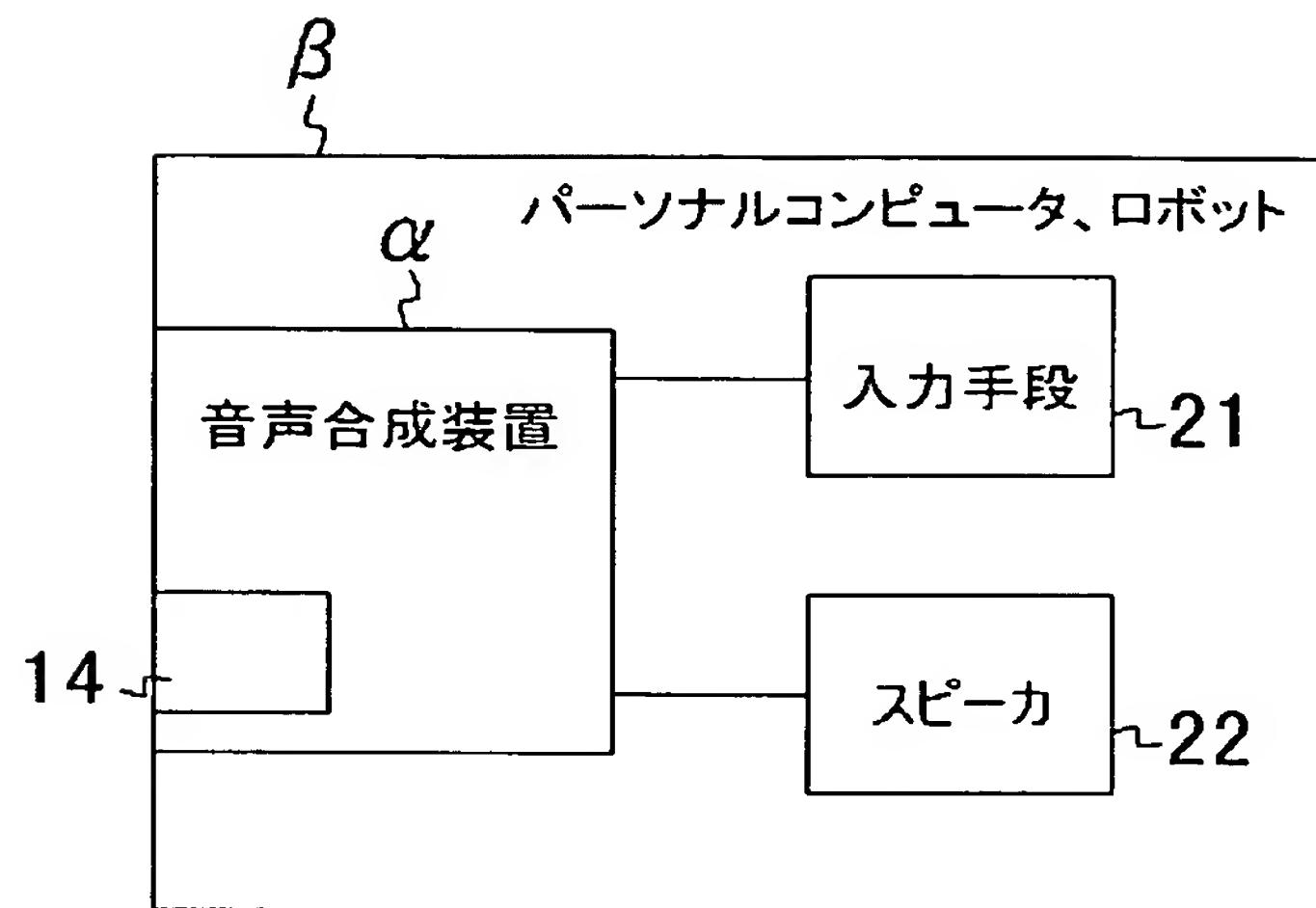
[図5]



[図6]



[図7]



THIS PAGE BLANK (USPTO)